EXChange AI · Modul 3

KI und Gesellschaft

Lehrnotizen für Erwachsenenbildner · Projekt **Expeditions for Change: Navigating the Skills & AI Landscape** · Projektnummer 2024-2-BG01-KA210-ADU-000296153 · exchangeai.eu

Co-funded by the European Union

Das Projekt wird von der Europäischen Union finanziert. Die geäußerten Ansichten sind die der Autorinnen und Autoren und spiegeln nicht notwendigerweise die Position der EU oder der Förderbehörde wider.

Hinweise zur Verwendung

Zweck dieser Notizen

Diese Notizen sind so geschrieben, dass sie direkt in das PowerPoint-Feld **Notes** zum jeweiligen Slide kopiert werden können. Der Text dient als **Lehrskript**: was erklärt werden soll, wie der Übergang zum Thema gestaltet wird und wie die Zielgruppe aktiviert werden kann.

Struktur des Moduls

01

Einführung & Goldilocks-Regel

Folien 1–3: Ausgewogenes Denken über KI

02

Diskriminierung & Vorurteile

Folien 4–7: Bias in KI-Systemen

03

Adversariale Angriffe

Folien 8–12: Sicherheit und Robustheit

04

Missbrauch, Arbeit & Planet

Folien 13–21: Gesellschaftliche Folgen

Folien 1–2: Einstieg & die Goldilocks-Regel

Folie 1 · Moduleinführung

Eröffnen Sie das Modul, indem Sie das Thema vorstellen: Es geht nicht nur darum, **was** künstliche Intelligenz ist, sondern auch darum, **wie** sie bereits Gesellschaft, Arbeit, Bildung, Vertrauen und Entscheidungsprozesse beeinflusst.

Betonen Sie, dass das Ziel ein **realistischer Blick** ist: weder Werbung für KI noch ein Angstszenario, sondern ein ausgewogenes Verständnis von Nutzen, Grenzen und Risiken.

Verbinden Sie das Thema mit den Bedürfnissen erwachsener Lernender: KI-Kompetenz bedeutet mehr als Werkzeuge bedienen — sie umfasst das Verstehen gesellschaftlicher Folgen.

Folie 2 · Die Goldilocks-Regel

Führen Sie die Goldilocks-Regel als Metapher für ausgewogenes Denken ein. Die öffentliche Diskussion über KI bewegt sich häufig zwischen zwei Extremen:

Übertriebener Optimismus

KI wird alle Probleme lösen

Die dritte Position

KI ist ein mächtiges Werkzeug — kein Zaubermittel, kein moralischer Akteur

Übertriebener Pessimismus

KI führt zwangsläufig in die Katastrophe

Aktivierungsfrage: „Welche dieser drei Positionen begegnet Ihnen am häufigsten in den Medien oder unter Kolleginnen und Kollegen?"

Folie 3: Grenzen von KI-Systemen

Erklären Sie, dass jedes KI-System Grenzen hat, auch wenn die Ergebnisse beeindruckend wirken. Das Beispiel aus der **medizinischen Bilddiagnostik** zeigt die Stärke von KI bei der Mustererkennung — aber auch die Notwendigkeit sorgfältiger Interpretation.

Leistungsfähigkeit

Ein System kann unter Testbedingungen gut funktionieren, aber bei anderen Daten, anderen Patientinnen und Patienten oder einem anderen Kontext **Fehler machen**.

Erklärbarkeit

Manchmal ist schwer nachvollziehbar, **warum** ein Modell genau dieses Ergebnis liefert. In sensiblen Bereichen reicht eine Antwort allein nicht aus.

Menschliches Urteil

KI sollte professionelle Urteilsfähigkeit **unterstützen**, aber nicht unkritisch ersetzen — besonders in Medizin, Finanzen und Bildung.

Folien 4–7: Diskriminierung und Vorurteile

Folie 4 · Einführung Bias

KI entsteht nicht in einem leeren Raum. Sie wird mit Daten trainiert, die von Menschen und Gesellschaften erzeugt wurden — und diese Daten enthalten oft **historische Ungleichheiten, Stereotype und unvollständige Repräsentation** verschiedener Gruppen.

⚠ Bereiten Sie die Teilnehmenden darauf vor, Vorurteile nicht als bloßen „Computerfehler" zu betrachten, sondern als gesellschaftliches Problem, das durch Automatisierung **verstärkt** werden kann.

Besonders für Lehrende wichtig: Bildung sollte Fairness, Inklusion und kritisches Denken fördern.

Folie 5 · Stereotype in Sprache und Daten

KI kann Stereotype aus Sprache und Daten lernen. Beispiel aus Wortvektoren:

Mann → Programmierer

Häufige Assoziation in Trainingsdaten

Frau → Hausfrau

Nicht wahr oder wünschenswert — aber statistisch häufig

Das Modell „versteht" Fairness nicht; es erkennt **statistische Muster**. Deshalb sind menschliche Überprüfung, Datendesign und ethische Aufsicht so wichtig.

Aktivierung: Bitten Sie Teilnehmende, Beispiele für Stereotype in beruflicher oder pädagogischer Sprache zu nennen.

Folien 6–7: Reale Folgen und Gegenmaßnahmen

Folie 6 · Reale Folgen von Bias

Vorurteile sind relevant, weil KI-Systeme zunehmend in **Entscheidungen mit hoher Tragweite** eingesetzt werden:

→ Bewerberauswahl

Automatisierte Filterung von Lebensläufen

→ Gesichtserkennung

Höhere Fehlerquoten bei bestimmten Gruppen

→ Kreditvergabe & Risikobewertung

Systematische Benachteiligung möglich

→ Personalisierung

Verstärkung bestehender Ungleichheiten

⊗ Der Schaden ist nicht nur individuell, sondern **systemisch**: Bei massenhafter Anwendung können ganze Gruppen benachteiligt werden.

Folie 7 · Gegenmaßnahmen

Die Bekämpfung von Vorurteilen erfordert eine Kombination aus technischen, organisatorischen und sozialen Maßnahmen:

Technisch

Reduzierung von Verzerrungen in Wortvektoren, sorgfältigere Datenauswahl, Tests mit verschiedenen Gruppen, regelmäßige Audits

Organisatorisch

Transparenz, klare Verantwortlichkeiten, diverse Entwicklungsteams

Pädagogisch

Ergebnisse prüfen, Perspektiven vergleichen, automatische Antworten nicht als neutrale Wahrheit darstellen

Folien 8–12: Adversariale Angriffe auf KI

Es geht hier nicht um gewöhnliche Fehler, sondern um **absichtliche kleine Veränderungen** in Daten oder Umgebung, die ein Modell täuschen können. Für Menschen kann die Veränderung fast unsichtbar sein — für das System führt sie zu einer völlig anderen Klassifikation.

Folie 9 · Visuelle Logik

Ein kleines Störsignal wird einem Bild hinzugefügt — das Modell ändert seine Antwort dramatisch (z. B. von „Kolibri" zu „Hammer"). Für das menschliche Auge bleibt das Bild fast gleich, aber die mathematische Repräsentation verändert sich stark genug, um das Modell zu verwirren.

Folie 10 · Physische Angriffe

Adversariale Angriffe können physische Formen annehmen: Brillen, Aufkleber, Pflaster, veränderte Verkehrsschilder. Ein Modell kann eine Person falsch erkennen oder ein Stoppschild übersehen.

Folie 11 · Selbstsichere Fehler

Systeme sagen häufig nicht „Ich weiß es nicht", sondern geben eine **selbstsichere, aber falsche** Antwort. Das ist auch eine wichtige Lektion für generative KI: Die Sicherheit der Ausgabe ist kein Beweis für Wahrheit.

Folie 12 · Schutzmaßnahmen

Training mit adversarialen Beispielen, strengere Tests, Monitoring, Filterung von Eingaben. Jede Schutzmaßnahme hat Kosten — und wie bei Spam und Anti-Spam passen sich Angreifer an. KI-Sicherheit ist ein **kontinuierlicher Prozess**.

Aktivierungsfrage (Folie 10): „Wo wären wir über eine solche Schwachstelle am meisten besorgt — in sozialen Medien, in Autos, in Banken oder in Schulen?"

Folien 13–14: Missbrauch von KI

Folie 13 · Einführung Missbrauch

Unterscheiden Sie zwischen **unbeabsichtigtem Schaden** und **absichtlichem Missbrauch**. Hier geht es darum, wie Menschen KI gezielt für Betrug, Manipulation, Überwachung oder Desinformation einsetzen können.

KI ist nicht „schlecht" — aber mächtige Werkzeuge können für unterschiedliche Zwecke genutzt werden. KI-Kompetenz umfasst daher auch die Fähigkeit, Inhalte zu **erkennen, zu überprüfen und kritisch zu bewerten**.

Folie 14 · Konkrete Missbrauchsformen

Deepfakes

Video oder Audio, in denen eine Person scheinbar etwas sagt oder tut, was sie nie gesagt oder getan hat — untergräbt Vertrauen in Medien und Institutionen.

Massenhafte Falschkommentare

Erzeugen eine Illusion öffentlicher Meinung durch automatisierte, koordinierte Inhalte.

Automatisierter Betrug

Überzeugendere und personalisiertere Betrugsversuche durch KI-Generierung.

KI-gestützte Überwachung

Gesichtserkennung und Verhaltensanalyse können in repressiven Kontexten Freiheiten und Rechte einschränken.

ⓘ Lehrende sollten Medienkompetenz einbauen: Quellen prüfen, Kontext suchen, manipulative Signale erkennen und Inhalte vorsichtig teilen.

Folien 15–18: KI, Arbeit und Beschäftigung

KI ist nicht nur eine technologische Frage — sie hat **wirtschaftliche Dimensionen**. Einerseits kann sie Produktivität steigern und neue Berufe schaffen. Andererseits kann sie die Struktur von Arbeit verändern und neue Kompetenzen erfordern.

Folien 16–17 · Automatisierungspotenzial

Stellen Sie die Zahlen als **Szenarien** dar, nicht als exakte Prognose. Automatisierung bis 2030 kann Hunderte Millionen Arbeitsplätze betreffen, während gleichzeitig neue entstehen sollen. Wichtig ist, nicht nur auf „Ersetzung" zu schauen, sondern auf **Transformation**.

Büroarbeit

Hohes Automatisierungspotenzial

Produktion

Erhebliche Veränderungen

Kundendienst

Generative KI erhöht Potenzial

Rechtliche Tätigkeiten

Früher schwerer automatisierbar

Aktivierungsfrage: „Welche Aufgaben in Ihrer Arbeit können durch KI unterstützt werden, und welche sollten unter menschlicher Kontrolle bleiben?"

Folie 18 · Lösungsansätze

Lebenslanges Lernen

Menschen können sich nicht nur auf Wissen aus der Erstausbildung verlassen. Kontinuierliche Kompetenzaktualisierung ist zentral.

Soziale Absicherung

Bedingtes Grundeinkommen als Beispiel: Sicherheitsnetz verbunden mit Anreizen für Lernen und Umschulung.

Rolle der Lehrenden

Erwachsenenbildner vermitteln zwischen technologischem Wandel und den realen Bedürfnissen der Menschen.

Folien 19–20: KI und ökologischer Fußabdruck

Folie 19 · Lebenszyklus großer KI-Modelle

Große Modelle haben einen Lebenszyklus mit unterschiedlichem Fußabdruck in Bezug auf Emissionen, Wasser und Materialien:

1 Konzeption & Entwicklung

Planung und Architektur des Modells

2 Bau von Rechenzentren

Infrastruktur und Hardwareproduktion

3 Training des Modells

Hoher Energieverbrauch — Standort und Energiemix entscheidend

4 Nutzung & Netzwerkverkehr

Jede Anfrage hat einen realen Fußabdruck

ⓘ Digitale Dienste sind nicht „gewichtslos" — sie nutzen Strom, Wasser, Materialien und Infrastruktur.

Folie 20 · Bewusste Nutzung

Diese Folie macht Umweltauswirkungen konkreter: Auch eine einzelne Anfrage hat einen kleinen, aber realen Fußabdruck. Wichtig ist, nicht zu dramatisieren, sondern eine **Gewohnheit vernünftiger Nutzung** zu entwickeln.

Praktischer Rat für Lehrende: Teilnehmende in effizienter KI-Nutzung schulen:

- **Klare Prompts formulieren**

  Präzise Anfragen reduzieren unnötige Iterationen

- **Fragen bündeln**

  Mehrere Anliegen in einer Anfrage zusammenfassen

- **Passendes Werkzeug wählen**

  Große Modelle nicht für triviale Aufgaben einsetzen

- **Vor dem Generieren prüfen**

  Unnötige Wiederholungen vermeiden

> So verbinden wir Produktivität mit Verantwortung.

Folie 21: Modulabschluss & Zusammenfassung

KI-Kompetenz ist nicht nur eine technische Fähigkeit — sie ist eine bürgerschaftliche, berufliche und ethische Kompetenz.

Vorurteile & Diskriminierung

Bias in Daten und Systemen erkennen, hinterfragen und bekämpfen

Adversariale Angriffe & Sicherheit

Robustheit von KI-Systemen verstehen und kritisch bewerten

Missbrauch & Desinformation

Inhalte erkennen, überprüfen und Medienkompetenz fördern

Arbeit & Planet

Wirtschaftliche und ökologische Dimensionen von KI verantwortungsvoll einbeziehen

Ermutigen Sie Lehrende, KI-Kompetenz durch **Diskussionen, praktische Aufgaben, Fallanalysen und kritische Reflexion** in ihre Kurse einzuführen.

exchangeai.eu besuchen